Azim Premji University

CSE Working Paper

#48

Did the nation-wide implementation of e-FMS in MGNREGS result in reduced expenditures? A re-examination of the evidence

Deepti Goel, J.V. Meenakshi and Zaeen De Souza

August 2022

Centre for Sustainable Employment

cse@azimpremjiuniversity.edu.in

Did the nation-wide implementation of e-FMS in MGNREGS result in reduced expenditures? A re-examination of the evidence

Deepti Goel[1], J. V. Meenakshi[2], and Zaeen De Souza[1]

[1]Azim Premji University
[2]Delhi School of Economics

Abstract

This paper revisits a part of the analysis by Banerjee et al. (2020), in which they examine the consequences of the nation-wide scale up of reforms to the funds management system (e-FMS) in India's national workfare programme, using a two-way fixed effects specification. They report a substantial 19 percent reduction in labour expenditures. We exploit the recent literature that highlights the limitations of the TWFE estimator in the presence of staggered roll out and effect a Goodman-Bacon decomposition of the TWFE coefficient, to pinpoint sources of identifying variation. We undertake a detailed examination of subsamples of six constituent and valid DiDs based on timing of treatment that are averaged into the TWFE coefficient to identify heterogeneity in treatment effects. This disaggregated subsample analysis does not support the conclusion of any reductions in MGNREGS labour expenditures, suggesting that the TWFE coefficient based on the full sample is indeed biased.

Keywords: e-governance, public funds management, MGNREGS

JEL classifications: H53, H75, D73, D78, I38

1 Introduction

There is an emerging literature that attempts to assess the development impacts of policy interventions implemented *at scale* to address concerns of external validity and to capture any general equilibrium effects that the intervention may have induced. Some of these integrate multiple sources of administrative and other secondary data, and use appropriate estimation strategies to make causal claims for large populations. Examples include Bharadwaj et al., 2020 and Gollin et al., 2021. Yet others employ large-scale (as opposed to pilot) randomized control trials (RCTs). In India, many of these have been effected at the state level, and have involved working with implementing agencies in designing a randomized roll out. An oft-cited example of such an evaluation is Banerjee et al., 2020.

Banerjee et al. (2020) examine whether e-governance can improve the functioning of the Mahatma Gandhi National Rural Employment Guarantee Scheme (MGNREGS), the national workfare programme in India.[1] They employ both approaches, that is, an RCT across all districts of Bihar, a large state in eastern India, as well as administrative data to analyse the nation-wide scale-up of the reform. The intervention basically involved a change to the process by which MGNREGS funds are transferred from the central government to the gram panchayats (GPs), the lowest level implementing agency operating at the level of a village or a group of villages. It consisted of moving from an advance-payment system,under which funds were transferred to the GP's account before work under the scheme was undertaken, to a 'just-in-time' payments system. Under the latter, funds were transferred soon after work was implemented, upon entering the work-related details in a central portal. The intent was to make it easier to detect fund leakages by reducing the time between actual work and the audit process, ultimately leading to savings in fiscal expenditure on account of reduced expenditures under MGNREGS. The main substance of Banerjee et al., 2020 analyses the impact of this reform on MGNREGS expenditures by employing a randomized roll out in Bihar. Much has been written about the experimental aspects of this and similar types of large scale evaluations (see Dreze, 2022 and Khera, 2021); but this is not the focus of our paper.

[1]The MGNREGS is India's flagship income support scheme that guarantees a 100 days of work to any rural household that demands it, in return for remuneration calculated at state minimum wages.

Instead, as the title of our note suggests, we re-examine the analysis that uses administrative data, presented in the last section (IV) of Banerjee et al., 2020 (henceforth BDIMP). This section is titled 'Did the Experimental Impacts Scale Up?' and examines the consequences for expenditures under MGNREGS arising from the *nation-wide* adoption of the electronic fund management system (e-FMS), which, at its core, involved the change to the fund transfer process described above. Between 2012 and 2015, the e-FMS reforms were progressively adopted across the country. In this part of their analysis, they rely on a two-way fixed effects (TWFE) specification to conclude that "e-FMS decreased expenditures by 19 percent....and the effect is persistent" (pp. 67-68).

In particular, we revisit this section in light of recent work in econometrics by Callaway and Sant'Anna (2021) and Goodman-Bacon (2021) that has highlighted the limitations of the TWFE method when analyzing interventions that involve variation in timing of implementation (as is the case here) and where there are time-varying impacts of the treatment. We attempt to unpack BDIMP's TWFE impact estimate into its constituent components, both to assess its validity, and to examine if there is impact heterogeneity depending on the characteristics of districts that adopted e-FMS first versus those that adopted it later. To effect this, we bring the Goodman-Bacon decomposition lens to bear on BDIMP's analysis of the impact of the national-level scale up of e-FMS on MGNREGS expenditure toward labour/wage payments.[2] We effectively re-estimate the upper panel of Tables 8 and Appendix Table A.17 of their paper, but disaggregate the TWFE impact estimate from the full sample into its constituents based on 12 canonical 2×2 DiD comparisons, 6 of which are invalid. We discard the 6 invalid comparisons and focus on analysing each episode of intervention separately, thus, overcoming the perils of using TWFE in the context of staggered interventions. We explain this in greater detail in section 3.

2 The full-sample TWFE specification

We begin by replicating BDIMP's TWFE specification (equation 5 of their paper). It takes the form:

$$Y_{dt} = \alpha + \beta EMFS_{dt} + \eta_d + \mu_t + \varepsilon_{dt} \tag{1}$$

[2]BDIMP also study the impact on material expenditure, but here, we only examine labour expenditure.

where the subscript d refers to district, and t to year, ranging from 2008 to 2016.[3] $EFMS_{dt}$ is an indicator for whether e-FMS (for labour expenditure) was operational or not. BDIMP find that the estimated β coefficient is ₹-770.4 lakh[4] (column 1 of Table 8 of BDIMP); this translates into a 18.6 (or approximately 19) percent reduction in labour expenditure (attributed to lower leakage) relative to a control group mean of ₹4140.4 lakh. In another specification they include introduction of e-FMS for material expenditure as a control and find that the result remains largely the same (column 2 of Table 8 of BDIMP).

Additionally, to examine whether programme effects persisted over time, they estimate the following (equation 6 of their paper):

$$Y_{dt} = \alpha + \beta EMFS_{dt} + \gamma EMFS_{d(t-1)} + \eta_d + \mu_t + \varepsilon_{dt} \tag{2}$$

where $EFMS_{d(t-1)}$ is an indicator for whether the e-FMS (for labour expenditure) was also operational in year $(t - 1)$. The γ coefficient is indicative of whether the e-FMS impact persisted over time. Once again, they examine whether coefficients are robust to the inclusion of e-FMS for material expenditure. In both cases, they find that the e-FMS effect of a reduction in labour expenditure persists over time (columns 3 and 4 of Table 8 of BDIMP).

Using data downloaded from the *American Economic Journal: Applied Economics* website, we first successfully replicated all their estimates on their sample of 473 districts. However, in order to use the Goodman-Bacon decomposition on their β coefficient, it is necessary to use a balanced panel. We therefore re-estimate the above specifications, using instead a balanced panel of 471 districts across the same 9 years, thus dropping two districts (and 14 observations) from the original sample in the paper.[5]

Table 1 reproduces the upper panel of Table 8 of BDIMP except with the balanced panel of 471 districts and shows that this change to the estimation sample results in a lower TWFE coefficient of ₹-644 lakh (column 1). Relative to a control group mean in our estimation sample of ₹4031.2 lakh, this figure constitutes a somewhat smaller, 16 percent, but nevertheless statistically signifi-

[3]Since our focus is only on labour expenditure, we dispense with their j^{th} subscript.
[4]One lakh equals 100,000.
[5]The two districts are Bardhaman and Medinipur, both in West Bengal.

cant and substantive reduction in labour expenditure. The remaining coefficients across different specifications in columns (2) to (4) are also somewhat lower as a consequence of dropping the two districts, but are of the same order of magnitude as reported by BDIMP.

Table 1. **Effect of e-FMS adoption on labour expenditure: TWFE results**

	District expenditure on labour (₹100,000)			
	(1)	(2)	(3)	(4)
e-FMS for labour in year t	-644.0***	-594.5***	-582.9***	-653.4***
	(150.4)	(147.2)	(154.3)	(169.8)
e-FMS for material in year t		-173.2	-232.7	247.2
		(181.3)	(197.7)	(198.7)
e-FMS for labour in years t and t-1			128.7	156.8
			(176.6)	(175.2)
e-FMS for labour in years t, t-1 and t-2				-470.8**
				(231.7)

Note: Control group mean is ₹4031.2 lakhs. * p-value<0.1, ** p-value<0.05, *** p-value<0.01
Source: Estimated on a balanced panel of 471 districts using BDIMP's data.

3 Staggered treatment timing and the Goodman-Bacon decomposition

As noted earlier, the nation-wide roll out occurred in a staggered manner, so that different districts adopted the e-FMS (were treated) in different years. Most of the variation in the year of adoption is at the state level (see Appendix Table A.16 of BDIMP's paper). Nevertheless, we use district-level variation in treatment timing, in keeping with their specification.

Considering the balanced panel of 471 districts, Table 2, column (1) shows that prior to 2012, the e-FMS (for labour expenditure) had not been adopted in any district. The majority of districts adopted e-FMS in 2012 and 2013, and by 2015 all 471 districts came under the ambit of e-FMS. Thus, 2008-2011 are the pre-treatment years. Furthermore, there are no "never treated" nor "alwaystreated" districts; and once treated, a district remained treated; in other words, there is no attrition from treatment.

Table 2. **Characterising districts by timing of e-FMS adoption**

Year e-FMS adoption (1)	# districts adopting (2)	% high leakage districts (3)	% rural households with mobiles (4)	% rural households with electricity (5)	% rural literacy (6)
2012	136	66	46	65	59
2013	276	47	50	57	61
2014	41	39	44	23	53
2015	18	38	57	47	58
Total	471	52	48	57	59

Source: Information on timing of adoption and characterisation of high leakage districts are compiled from the BDIMP's data. Data on access to mobiles, electricity and literacy are compiled from unit record data of the 2011 census. All figures refer only to rural areas/households.

Given this staggered nature of the nation-wide roll out of e-FMS implementation, it is possible that the 19 (16, in the balanced panel) percent reduction in labour expenditure is in part driven by certain "forbidden" comparisons (explained below). It is also quite likely that it subsumes substantial heterogeneity. Using the Bacon-Decomposition we try and establish if these possibilities bear out.

In the standard 2×2 canonical version of the Difference in Differences (DiD) estimator, the identifying assumption is simply that, in the absence of intervention, treated units would have evolved in the same way as the observed evolution of the untreated units. If sufficient pre-treatment periods are available, it is common to present evidence of parallel trends in the outcome between treated and untreated groups in the pre-intervention period as being suggestive that the identifying assumption holds.[6] Unlike the canonical DiD estimator, in the TWFE version of the DiD estimator, when there is staggered roll out of the intervention, identification not only requires parallel trends, but also that the treatment effects take the form of one-time level-shifts, devoid of any time varying effects.

In particular, Goodman-Bacon (2021) demonstrates that in a situation of staggered treatment timing with time varying effects, the TWFE estimator is a weighted average of effects emerging from

[6] While it is of course not a guarantee that parallel trends would also hold in the intervention period (which is what the identifying assumption pertains to), it does make it more plausible.

a series of 2×2 DiD comparisons: for k treatment timings, there are $(k^2 - k)$ 2-period 2-group DiD impact estimates embedded in the single TWFE estimate. Some of these 2×2 estimates are derived from inappropriate comparisons, in that they involve comparing units exposed to later treatment episodes with earlier treated units acting as the comparison group. The earlier treated units form an invalid control group whenever their outcome is changing on account of the treatment itself; in such a scenario they do not provide the correct counterfactual of what the later treated units would look like in the absence of the treatment (see Figure 3 of Goodman-Bacon's paper for more on the intuition). Yet others are valid 2×2 DiD estimates, but are derived using different pairs of treated and comparison groups, depending on when they were treated. It would, therefore, not be surprising if these estimates widely differed from each other. In fact, studying these differences may be of interest in order to investigate and characterize heterogeneity in treatment effects. The Goodman-Bacon decomposition of the TWFE estimate into its constituent estimates, plotted against the weights associated with each, provides a simple visual assessment of the sources of identifying variation, and more specifically tells us how much of the effect size is being driven by variation coming from comparisons that are forbidden.

For e-FMS roll-out, there are four episodes of treatment, or "timing groups", to use Goodman-Bacon's terminology, namely, 2012, 2013, 2014 and 2015. This in turn implies that the TWFE coefficient can be expressed as a weighted average of twelve different 2×2 DiD coefficients, six of which are readily interpretable, as they compare earlier-treated districts with sets of districts that are yet to be treated, and constitute "early vs. late" comparisons, in Goodman-Bacon's terminology. Thus, in one 2×2 DiD, districts treated in 2013 can serve as controls for those treated in 2012, as can districts treated in 2014 and 2015 in two other distinct DiDs. Similarly, districts treated in 2014 and 2015 can each serve as a comparison group for those treated in 2013; and districts treated in 2015 can serve as a comparison for districts treated in 2014, making a total of six 2×2 DiDs.

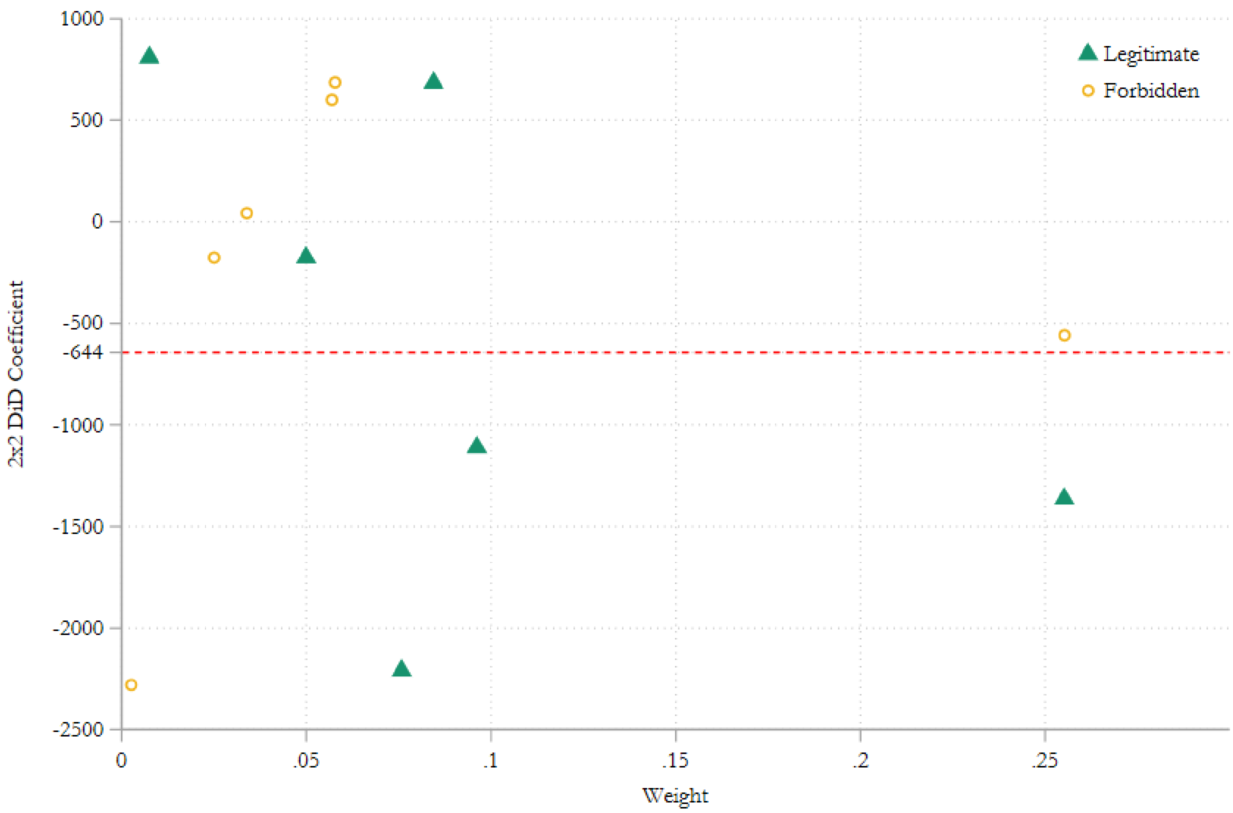
As Goodman-Bacon points out, there are an additional six 2×2 DiD comparisons that produce estimates that are averaged into the TWFE coefficient and are characterised as "late vs. early" comparisons. These are invalid, since they involve comparison units that have already been treated in an earlier period. Even though, once treated, their treatment status does not change subsequently, their outcomes may still be changing on account of the earlier treatment, rendering them ineffective as controls.

Note further that the e-FMS roll out was not random in nature. Table 1 also presents some district characteristics such as literacy,[7] culled out from the 2011 census for each of the four timing groups. In addition, we examined if access to rural e-infrastructure played a role in determining where e-FMS was rolled out first. We use the percentage of rural households with electricity and the percentage of rural households reporting having at least one mobile phone, as proxies for availability of e-infrastructure. Apart from these, we also include a binary variable constructed by BDIMP who characterised a district as being high leakage or not. Districts with above national median leakage were characterised as high leakage (coded 1), and the rest as low leakage (coded 0). Districts that adopted the e-FMS in 2012 were far more likely to have been classified as high leakage than those treated in 2013 or later (column 2, Table 2). They were also more likely to have had a higher proportion of rural households with electricity (column 3) and mobile phone connectivity (column 4) and be more literate (column 5). While it makes sense for the e-FMS reforms to have been targeted to districts that had higher levels of leakage, and had better e-infrastructure, this also implies that impacts may vary based on when districts received the treatment, and the nature of the comparison groups being used to calculate impact.

Figure 1 presents Goodman-Bacon decomposition for the baseline specification (corresponding to column 1 of Table 2), plotting each 2×2 DiD coefficient against its weight in the overall TWFE coefficient of -644. Appendix Table A1 presents the same information in tabular form. As anticipated, there is a wide range in the magnitudes of the 2×2 coefficients that make up the TWFE coefficient. Of the six legitmate (early vs. late, indicated by triangles) canonical DiDs, 2 are perverse (positive, indicating that expenditures increased as a consequence of the adopting the e-FMS) in sign, and at least one of the forbidden (late vs. early, indicated by circles) DiD coefficients has a large weight in the TWFE. When we consider only the six legitimate DiD coefficients, and compute the corresponding TWFE coefficient using rescaled weights, we get a figure of -996.7, implying a much larger reduction in expenditure, of almost 25 percent relative to our control group mean.

[7] We also looked at scheduled caste and scheduled tribe populations, but do not present these in the interest of space.

Figure 1 **Goodman-Bacon decomposition of the TWFE coefficient**



Note: See Appendix Table A1

4 Impact heterogeneity in national scale-up of e-FMS?

Next, in our attempt to investigate the heterogeneous effects seen above, we delve further into the subset of six coefficients that represent legitimate comparisons (canonical DiDs) in their own right, and present them along with their corresponding standard errors. If there is substantial heterogeneity in impacts, emanating from when and where treatment is rolled out, the magnitude of the coefficients would vary substantially, while coefficients similar in magnitude would be indicative of relatively homogeneous treatment effects.

To establish this, we re-estimate each of the main specifications undertaken by BDIMP (and presented in Table 1 for the balanced sample) but disaggregated by the six subsamples corresponding to each of the legitmate early vs. late comparisons. Each subsample represents a *single* treatment-timing vs untreated group comparison. We exclude the six forbidden comparisons that are also embedded in the full sample TWFE coefficients. This disaggregated look enables us to unpack any treatment heterogeneity by timing group.

Table 3 presents a disaggregated version of the estimations in Table 1. Panel A presents the 2×2 canonical DiDs for the districts treated in 2012, compared to the comparison group consisting of districts that were treated later in 2013. Thus, for this DiD only years 2008-2012 are considered. The other panels are similarly defined. Column (1) provides the 2×2 DiD coefficient corresponding to the baseline specification with no control. A comparison across the panels suggests that only three of the six constituent legitimate DiD coefficients are negative and statistically significant; one is negative but insignificant, and the other two are, in fact, perverse (also seen in Figure 1), but also insignificant. At first glance the coefficients in column (1) in Panels A, B and D are seemingly indicative of substantial heterogeneity in treatment effects, with much larger effect sizes in some subsamples than the 19 (16) percent reported in Banerjee et al. (we find). They amount to a reduction in labour expenditure (and by implication leakages) to the tune of between 28 percent (-1110) and 55 percent (-2209) of ₹4031.2 lakh, the mean labour expenditure in the control group for our balanced panel.

Table 3. **Effect of e-FMS adoption on labour expenditure**
2 × 2 canonical DiD estimates presented separately

	District expenditure on labour (₹100,000)			
	(1)	(2)	(3)	(4)
Panel A: (Treated group 2012, Comparison group 2013)				
e-FMS for labour in year t	-1363.4***	-1368.3***		
	(272.5)	(273.6)		
Panel B: (Treated group 2012, Comparison group 2014)				
e-FMS for labour in year t	-2209.3***	-2340.2***	-2344.4***	
	(368.2)	(461.6)	(476.6)	
e-FMS for labour in years t and t-1			46.1	
			(1289.4)	
Panel C: (Treated group 2012, Comparison group 2015)				
e-FMS for labour in year t	-175.3	-188.2	-563.2	-566
	(419.1)	(529.9)	(613.9)	(614.2)
e-FMS for labour in years t and t-1			1130.8	785.1
			(935.2)	(1101.5)
e-FMS for labour in years t, t-1 and t-2				478.9
				(805.3)
e-FMS for material in year t	N	Y	Y	Y

See next page for continuation of Table 3.

Column (2) presents the same specification, but with the addition of e-FMS for material as an additional control. In this specification, however, only two of the DiD coefficients, corresponding to Panels A and B retain their expected sign and significance. In Panel E, the effect is significant and is perverse, suggesting relative to group treated in 2015, those districts that adopted e-FMS in 2013 had much higher expenditure on labour.

Columns (3) and (4) examine if there is persistence in effects across years, again mimicking the same columns of Table 8 in BDIMP's paper. Estimation including $EFMS_{d(t-1)}$ term is feasible only for three timing groups: namely those corresponding to Panels B, C and E, while inclusion of both $EFMS_{d(t-1)}$ and $EFMS_{d(t-2)}$ terms is feasible only for one group corresponding to Panel C. None of the estimated γ coefficients are significant in these subsamples, implying no evidence of effects increasing nor decreasing over time. In fact, for these three subsamples, only one (Panel

Table 3 continued. **Effect of e-FMS adoption on labour expenditure**

	District expenditures on wages (₹100,000)			
	(1)	(2)	(3)	(4)
Panel D: (Treated group 2013, Comparison group 2014)				
e-FMS for labour in year t	-1110.6***	97.6		
	(428.1)	(490.9)		
Panel E: (Treated group 2013, Comparison group 2015)				
e-FMS for labour in year t	683.4	1681.0***	1339.0**	
	(462.4)	(511.8)	(643.3)	
e-FMS for labour in years t and t-1			682.5	
			(777.5)	
Panel F: (Treated group 2014, Comparison group 2015)				
e-FMS for labour in year t	808.7	732.1		
	(636.4)	(664.3)		
e-FMS for material in year t	N	Y	Y	Y

Note: All specifications include district and year fixed effects. * p-value<0.1, ** p-value<0.05, *** p-value<0.01
Source: Estimated using BDIMP's data.

B) of the estimated βs has the expected negative sign and is significant. In the other two cases it is either insignificant (Panel C) or perverse (Panel E). Furthermore, in Panel D, the sign on the estimated β flips across the two specifications, from being negative and significant in column (1) to positive, although insignificant, in column (2). Finally, the perverse sign in Panel E persists across specifications. It is only in Panels A and B that the negative and significant magnitudes of the canonical DiD coefficients are robust across specifications.

Given the wide variation in the coefficients across subsamples and specifications, we examine if there are differential trends in the pre-implementation period to see if the identifying assumption of parallel trends is validated. This is straightforward to do, since, by construction, each panel is characterised by a single timing of treatment. Corresponding to each, we can define a pre-treatment period. Thus, for example, for districts treated in 2012, 2008-2011 serves as the pre-treatment period, for districts treated in 2013, 2008-2012 constitute the pre-treatment period. Therefore as a falsification check, for each legitimate DiD, we only considered data from its pre-treatment period and defined a placebo treatment in 2010 in all six cases. An insignificant estimated β would support (but not ensure) the identifying assumption of parallel trends. We implement this for the first two specifications corresponding to columns (1) and (2) and find, however, that in five of the six cases, there were significant differences in pre-intervention trends between treated and comparison districts; the exception being the (2013, 2015) group in Panel E.[8] Thus, most of the results from even the canonical DiD specifications are untenable, given strong evidence of differing trends between control and treated groups during the pre-implementation period. For the one (2013, 2015) comparison for which parallel trends are met, the estimated β are perverse and the magnitudes vary dramatically across specifications, rendering it also unreliable.

These results are strongly indicative of differential trends across districts prior to implementation. To account for this possibility, BDIMP include district-specific time trends, and find that it does not alter the substance of their conclusion of significant and persistent savings from the nation-wide adoption of e-FMS. Like them, we also find that when district-specific time trends are included in the *full* sample, the results are largely in line with those seen in Table 1. Appendix Table A2 presents the counterpart (based on 471 districts) to their Appendix Table A.17 (based on 473 districts).

But in a manner parallel to that set out above, when disaggregated across the six early vs. late subsamples, the inclusion of district-specific time trends in all four specifications renders nearly all the impact coefficients insignificant. Table 4 demonstrates.

None of the impact coefficients in column (1) of Table 4 are significant, in sharp contrast to what was seen for at least Panels A and B in Table 3. Even if one were to argue that insignificance is driven by inadequate sample sizes (although formal calculations for this are not possible), two of the six coefficients are perverse, and this includes the one in Panel A. Similarly, in column (2), in the specification that includes e-FMS for material as a control, five of six coefficients are insignificant; the remaining one (Panel D) is negative (-668.3), but only weakly significant at 10 percent, and the two that were perverse in column (1) continue to remain so. The addition of "lagged" terms

[8]These results are available with the authors on request. These results are in contrast to insignificance of coefficients during the pre-intervention phase that the authors report in the event study specification presented in their Figure 4; and are driven entirely by the use of subsamples (in our case) and the full sample (in theirs).

Table 4. **Effect of e-FMS adoption on labour expenditure**
2×2 **canonical DiD estimates presented separately**
Includes district-specific time trends

	District expenditure on wages (₹100,000)			
	(1)	(2)	(3)	(4)
Panel A: (Treated group 2012, Comparison group 2013)				
e-FMS for labour in year t	295.4	308.0		
	(247.9)	(248.9)		
Panel B: (Treated group 2012, Comparison group 2014)				
e-FMS for labour in year t	-231.7	-435.5	-460.1	
	(474.3)	(484.3)	(484.9)	
e-FMS for labour in years t and t-1			1078.1	
			(1062.8)	
Panel C: (Treated group 2012, Comparison group 2015)				
e-FMS for labour in year t	482.4	410.3	349.3	543.2
	(647.1)	(656.3)	(655.6)	(668.3)
e-FMS for labour in years t and t-1			1704.6**	1102.4
			(827.4)	(923.1)
e-FMS for labour in years t, t-1 and t-2				958.6
				(653.5)
e-FMS for material in year t	N	Y	Y	Y

See next page for continuation of Table 4.

to indicate if the e-FMS was available in years t-1 and t-2 in the specifications in columns (3) and (4) do not alter the basic lack of significance across the six DiD groups. If anything, adding the coefficients on the lagged terms to the main β coefficient, once again gives a perverse sign in all cases. In short, the addition of district-specific time trends to the underlying specifications lend support to our conclusion that there is no evidence of any reduction in expenditure on labour as a consequence of the adoption of the e-FMS.

5 Conclusions

BDIMP's estimate of a 19 percent reduction in MGNREGS expenditure on labour is, in principle, a major finding, irrespective of whether it can be attributed to lower leakage in the system. It is

Table 4 continued. **Effect of e-FMS adoption on labour expenditure Includes district-specific time trends**

	District expenditure on wages (₹100,000)			
	(1)	(2)	(3)	(4)
Panel D: (Treated group 2013, Comparison group 2014)				
e-FMS for labour in year t	-393.4	-668.3*		
	(335.4)	(386.8)		
Panel E: (Treated group 2013, Comparison group 2015)				
e-FMS for labour in year t	-321.0	-193.0	-336.1	
	(491.4)	(518.1)	(546.8)	
e-FMS for labour in years t and t-1			427.5	
			(521.4)	
Panel F: (Treated group 2014, Comparison group 2015)				
e-FMS for labour in year t	-925.6	-752.1		
	(587.4)	(612.5)		
e-FMS for material in year t	N	Y	Y	Y

Note: All specifications include district, year fixed effects and district-specific time trends.
Source: Estimated using BDIMP's data.

all the more remarkable since this number is in line with the results from the closely-monitored randomised experiment in the state of Bihar. It would seem that the nation-wide reform to the fund transfer systems between the Central government and the local governments that implement MGNREGS at the lowest level resulted in substantial savings to the national exchequer.

It is these substantive results, based on the TWFE estimation strategy, that motivated the present exercise. Our hope was to provide an adjusted estimate that overcomes the limitations of the TWFE estimation strategy by exploiting the relatively new literature that provides a correction to the TWFE estimator including, in addition, Sun and Abraham (2021) and Callaway and Sant'Anna (2021). Our hope was also to uncover heterogeneity in the effects of the e-FMS, by focusing on each of the valid constitutent canonical 2×2 DiDs. We then wanted to examine whether any heterogeneity in treatment effects was at all related to the differences in the characteristics of the districts that differed based on when they adopted the e-FMS.

Replicating the main specifications presented by BDIMP for each subsample of comparison groups, we find that the canonical DiD coefficients for two of the comparison groups (Panels A and B of Table 3, when much of the country adopted the e-FMS) do vary substantially. However, these (and other) comparison groups fail a basic falsification test, which indicates that trends in the outcome varied substantially across each treatment and control group in the pre-implementation period. It is thus hard to make the case for the identifying assumption that, absent the e-FMS, the MGNREGS expenditure in treated districts would have evolved in the same way as that in the comparison districts. This means that the finding of effect sizes between 28 to 55 percent (Table 3 of our paper) cannot be vested with causal meaning. Given the apparent failure of the identifying assumption in the canonical specification, we attempted alternative specifications that included district-specific time trends. Unfortunately, these yield effect sizes that are insignificant (and sometimes perverse), as noted in Table 4.

We are therefore forced to answer the question posed in the title of this paper in the negative, or at best, "unknown". Our results indicate that far from uncovering any heterogeneity in treatment effects, there is no evidence that the nation-wide scale up of the e-FMS produced any tenable reduction in MGNREGS expenditure on labour. This means there is no evidence to suggest that there was a reduction in leakages and thereby greater efficiency in MGNREGS implementation.

This note serves also as a cautionary note for empirical exercises that use the TWFE specification in the presence of staggered treatment timing, even when there is no attrition from the treatment. It is important to consider if identification conditions under staggered treatment timing, as outlined by Goodman-Bacon, and other literature on causal identification cited above, are satisfied, and to build up aggregate effects by examining legitimate subsamples of timing groups.

References

Banerjee, A., E. Duflo, C. Imbert, S. Mathew, and R. Pande (2020). E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in india. *American Economic Journal: Applied Economics 12*(4), 39–72.

Bharadwaj, P., J. Fenske, N. Kala, and R. A. Mirza (2020). The green revolution and infant mortality in india. *Journal of Health Economics 71*, 102314.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

Dreze, J. (2022). Perils of embedded experiments. *Ideas for India*.

Gollin, D., C. W. Hansen, and A. M. Wingender (2021). Two blades of grass: The impact of the green revolution. *Journal of Political Economy 129*(8), 2344–2384.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277.

Khera, R. (2021). Some questions of ethics in rcts. *Available at SSRN 3780908*.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics 225*(2), 175–199.

Appendix Tables

Table A1. Goodman-Bacon Decomposition of TWFE coefficient (-644.02)

Treated	Control	DiD Type	DiD Coefficient	Weight in TWFE
2012	2013	Early vs Late	-1363.35	0.255
2013	2012	Late vs Early	-557.24	0.255
2012	2014	Early vs Late	-2209.26	0.076
2014	2012	Late vs Early	601.44	0.057
2013	2014	Early vs Late	-1110.57	0.096
2014	2013	Late vs Early	687.83	0.058
2012	2015	Early vs Late	-175.28	0.050
2013	2015	Early vs Late	683.45	0.084
2014	2015	Early vs Late	808.69	0.008
2015	2012	Late vs Early	-174.56	0.025
2015	2013	Late vs Early	44.18	0.034
2015	2014	Late vs Early	-2278.00	0.003

Table A2. **Effect of e-FMS adoption on labour expenditure: TWFE results including district-specific time trends**

	District expenditures on wages (Rs.100,000)			
	(1)	(2)	(3)	(4)
e-FMS for labour in year t	-587.0***	-677.4***	-640.6***	-602.6***
	(162.8)	(161.3)	(173.9)	(198.2)
e-FMS for material in year t		331.9*	195.3	201.7
		(183.3)	(198.3)	(197.3)
e-FMS for labour in years t and t_{-1}			312.6	329.3
			(215.8)	(223.4)
e-FMS for labour in years t, t_{-1} and t_2				179.8
				(217.6)
District Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
District-specific Time Trend	Y	Y	Y	Y
Control Mean	4031.2	4031.2	4031.2	4031.2